APPENDIX: EXPERIMENTAL DETAILS

Dataset Description

The experiments were conducted on the following publicly available recommendation benchmarks:

MovieLens-1M[4]: It contains 1,000,209 user-system interactions from 6,040 users on 3,952 movies. We considered the user's occupation as the sensitive attribute, which contains 21 classes of occupations. For the user feedback, we treated the 5-star and 4-star ratings made by users as positive feedbacks (labeled with 1), and others as negative feedbacks (labeled with 0). For each user, we took the last interaction for model testing.

Insurance[5]: It is an insurance recommendation dataset on Kaggle, containing 5,382 interactions from 1,231 users on 21 insurances. We selected user's gender as the sensitive attribute that is a binary attribute, and split the dataset into a training set (80%), a validation set (10%) and a testing set (10%).

RentTheRunWay[6]: It contains 192,544 user-system interactions from 105,508 customers on 5,850 products, which were collected from a platform that allows women to rent clothes for various occasions. We selected the user's age as the sensitive attribute. We assigned users into 12 equal-length groups according to their age range. We also split the dataset into a training set (80%), a validation set (10%), and a testing set (10%).

Implementation Details

All the baselines and base models were trained on a single NVIDIA Tesla P100 GPU, with the batch size tuned among $\{64, 128, 256, 512, 1024\}$ and the learning rate tuned in {1E-1, 1E-2, 1E-3, 1E-4}. In the implementation of MACE, we set the networks g_{exo}, g_{endo}, MLP_1 and MLP_2 to 3-layer fully connected neural networks, respectively, where the activation functions were $\tanh$ and sigmoid. The dimensions of the exogenous part $g_{\text{exo}}(\mathbf{v}_u)$ and the endogenous part $g_{\text{endo}}(\mathbf{v}_u)$ were set to $d_{\text{exo}} = d_{\text{exo}} = 32$, and the dimensions of the user and item embeddings were all set to 128. In the training process of MACE, the batch size B and the maximum number of iteration N were set to 128 and $20 \times$ sample size$/128$, respectively, the update cycle ρ was set according to the sample size of training data ensuring that the MI minimization was executed 20 times, the fair weight λ in the final loss:

$$\mathcal{L}(\boldsymbol{\theta}_{\text{all}}) := \sum_{(\mathbf{s}_u, \mathbf{x}_u, \mathbf{v}_i, y_{u,i}) \in \mathcal{D}_{\text{train}}} \underbrace{\ell\left[\, f(h_{\text{re}}(\mathbf{v}_u), \mathbf{v}_i), y_{u,i}\right]}_{\text{Prediction Loss}} + \underbrace{\lambda \cdot \text{ECGF}(f, h_{\text{re}}, \mathcal{T})}_{\text{CGF-Oriented Constraint}} + \gamma \|\boldsymbol{\theta}_{\text{all}}\|_2^2,$$

was tuned among $[0 : +0.1 : 5]$, the regularization parameter γ in the final loss was set to 0.001, and the regularization parameter τ in the IV regression

$$\begin{aligned}\widehat{\boldsymbol{\Gamma}} &= \underset{\boldsymbol{\Gamma} \in \mathbb{R}^{d_{\text{endo}} \times d_{\text{exo}}}}{\arg\min} \sum_{\mathbf{v} \in \mathcal{V}_{\text{ue}}} \|g_{\text{endo}}(\mathbf{v}) - \boldsymbol{\Gamma} g_{\text{exo}}(\mathbf{v})\|_2^2 + \tau \|\boldsymbol{\Gamma}\|_{\text{F}}^2 \\ &= \mathbf{G}_{\text{endo}}^{(B)} \mathbf{G}_{\text{exo}}^{(B)\intercal} \left(\mathbf{G}_{\text{exo}}^{(B)} \mathbf{G}_{\text{exo}}^{(B)\intercal} + \tau \mathbf{I}_{d_{\text{exo}}} \right)^{-1},\end{aligned}$$

was set to 0.9.

For the baselines Mixup and MACE-GapReg, we chose Δ DP as the fairness constraint. For fair comparisons, in the CGF-oriented constraint of MACE, we chose a weighted version of Δ DP as the GF metric in ECGF, where the set of perturbation parameters in ECGF was set to $\mathcal{T} = \{0, 0.2, 0.4, 0.6, 0.8, 1\}$.

4. https://grouplens.org/datasets/movielens/1m/
5. https://www.kaggle.com/mrmorj/insurance-recommendation
6. https://www.kaggle.com/datasets/rmisra/clothing-fit-dataset-for-size-recommendation